INFSCI 1530 Data Mining

Department of Computer Science, SCUPI

Spring 2026

Instructor

Ziyi Wang

Email: ziyi.wang@scupi.cn

Office: N527

Office Hours: Friday 10:00-15:00 or by appointment

Lectures

- Section 1: Wednesday 08:15-11:00 @ S103
- Section 2: Wednesday 13:50-16:26 @ S103

Description

The ability to collect, store and process large amounts of detailed data in a variety of fields has led to a surge in the use of data in various decision-making tasks, ranging from governmental policy-making to drafting players in sports. Data literacy is thus important and in this first introductory course, we will focus on shifting the traditional mode of deterministic (yes/no) thinking to probabilistic thinking. We will see concepts from applied probability and statistics, while we will also explore a variety of data analysis methods including linear regression, matrix factorization, and network analysis.

Prerequisites

- INFSCI 0510 Data Analysis
- MATH 0280 Matrices and Linear Algebra

Course Objectives

- Explain the terminology and concepts of data mining
- Evaluate the possibilities and fundamental limitations of data mining
- Evaluate the relative advantages and disadvantages of major approaches to data mining
- Demonstrate the basic methods and techniques used in data mining
- Apply the basic methods and techniques to actual problems in data mining

Applicable ABET Outcomes

- Analyze a complex computing problem and apply principles of computing and other relevant disciplines to identify solutions.
- Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline.
- Communicate effectively in a variety of professional contexts.
- Recognize professional responsibilities and make informed judgments in computing practice based on legal and ethical principles.
- Function effectively as a member or leader of a team engaged in activities appropriate

to the program's discipline.
- Apply computer science theory and software development fundamentals to produce computing-based solutions.

Textbook
- Jiawei Han, Jian Pei, Hanghang Tong. *Data Mining: Concepts and Techniques, 4th Edition*. Morgan Kaufmann. 2023.

Reference
- Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar. *Introduction to Data Mining, 2nd Edition*. Pearson. 2018.
- Charu C. Aggarwal. *Data Mining: The Textbook*. Springer. 2015.
- Jure Leskovec, Anand Rajaraman, Jeff Ullman. *Mining of Massive Datasets, 3rd Edition*. Cambridge University Press. 2020.

Grading
- Attendance   10%
- Homework   40%
- Exam   50%

Conversion from Numerical Score to Letter Grade

100~90	89~85	84~80	79~76	75~73	72~70	69~66	65~63	62~61	60	<60
A	A–	B+	B	B–	C+	C	C–	D+	D	F

Communication
All lecture notes, learning materials, homework assignments, and announcements will be published on Blackboard (https://pibb.scu.edu.cn/). It is the student's responsibility to regularly check Blackboard in a timely manner. Important announcements and notifications will be sent to QQ group.

Class Policy
Attendance
Class attendance will be checked on a bi-weekly basis. A student has three chances to be absent from class free of penalty without prior notification. For one-time absence, if you have a strong and valid reason, please provide official proof to the instructor.

Homework
Homework assignments are submitted every four weeks. Late submission induces a penalty of 25% per week. For example, the first homework assignment is due at the end of week 4, submission at any time in week 5, 6, 7, and 8 induces a penalty of 25%, 50%, 75%, and 100%, respectively. Questions regarding the grading of homework assignment must be raised to the instructor within a one-week time frame.

Exam

There will be only one final exam. The exam will be OPEN BOOK, OPEN NOTES, and CLOSED COMPUTER.

Academic Integrity
The principles of academic integrity require that a student make sure that all work submitted is the student's own and created without the aid of impermissible technologies, materials, or collaborations. Academic integrity policy will be strictly followed.

Classroom Recording
To ensure the free and open discussion of ideas, students may not record classroom lectures, discussion and/or activities without the advance written permission of the instructor, and any such recording properly approved in advance can be used solely for the student's own private use.

Copyrighted Material
All material provided through this web site is subject to copyright. This applies to class notes, slides, assignments, solutions, project descriptions, etc. You are allowed (and expected!) to use all the material provided for personal use. However, you are strictly prohibited from sharing the material with others in general and from posting the material on the Web or other file sharing venues.

Tentative Course Schedule

Week	Topic
1	Introduction
2	Data, Measurements, and Data Preprocessing
3	Pattern Mining: Basics
4	Presentation
5	Pattern Mining: Advanced
6	Classification: Basics
7	Classification: Advanced
8	Presentation
9	Cluster Analysis: Basics
10	Cluster Analysis: Advanced
11	Outlier Detection
12	Presentation
13	Mining Text Data
14	Mining Time Series Data
15	Mining Web Data
16	Presentation
17	Final Exam